Chapter 6: Copying Files Remotely Using the dCache

If you are off-site and wish to use Enstore, you need to go through a dCache door[1]. You use ftp to copy files back and forth between your machine and your `/pnfs/storage-group` area on the machine running dCache. Enstore and dCache work together (in a manner transparent to the user) to transfer files between storage media and the disk caches on the dCache server. Fermilab's entire `/pnfs/` area is mounted on this server.

The dCache server node and the ports documented in this section are subject to change. You can always find the current ones from the web page `http://www-isd.fnal.gov/enstore/dcache_user_guide.html`.[2] Currently (May 2003), there are three dCache server nodes, each corresponding to an Enstore installation (STKEN, CDFEN, and D0EN). Each dCache server may have multiple doors, thus allowing a variety of access methods.

6.1 Simple Kerberized FTP

The dCache door for Kerberized ftp service enforces Kerberos authentication (see *Strong Authentication at Fermilab Documentation* at `http://www.fnal.gov/docs/strongauth/`). It currently runs on the following nodes and corresponding ports:

- fndca.fnal.gov, port 24127 (for STKEN)
- cdfdca, port 25127 (for CDFEN)
- d0endca, port 24127 (for D0EN)

(The port number is installation-specific.) Any Kerberized ftp client can be used on the client machine. You must specify the host port in your ftp command, as shown below.

Notes:

1. A dCache door is an instance of the dCache software installed on a server associated with a particular port and having its own access profile.
2. It is available from the *Fermilab Mass Storage Systems* home page (`http://hppc.fnal.gov/enstore/`); see the list of items under *Documentation* for dCache, and use the *User Access at FNAL* link.

- File reads and writes are supported when the user (a) is authorized by the experiment to access the data stores, and (b) has obtained Kerberos credentials.
- Portal Mode (CRYPTOCard) access is not supported. It is not compatible with automated transfers or future GRID development.

6.1.1 Sample Kerberized FTP session

User is authenticated to Kerberos and authorized for the Kerberized dCache door (currently at fndca.fnal.gov, port 24127):

```
% ftp fndca.fnal.gov 24127
```

```
Connected to stkendca3a.fnal.gov.
220 FTPDoorIM+GSS ready
334 ADAT must follow
GSSAPI accepted as authentication type
GSSAPI authentication succeeded
Name (fndca:aheavey):
200 User aheavey logged in
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> cd aheavey/test3
250 CWD command succcessful. New CWD is </aheavey/test3>
ftp> ls
200 PORT command successful
150 Opening ASCII data connection for file list
dupl2
duplexps
226 ASCII transfer complete
ftp> get duplexps
local: duplexps remote: duplexps
200 PORT command successful
150 Opening BINARY data connection for /pnfs/fs/usr/test/aheavey/test3/duplexps
226 Closing data connection, transfer successful
42 bytes received in 0.033 seconds (1.2 Kbytes/s)
ftp>
```

6.2 Kerberized ftp via the kftpcp Command

A regular ftp client (Kerberized or not) is an interactive program which is hard to use in batch mode. In order to access data from a batch job or a background process, you should either use ftp client libraries (available from many sources), or the **kftp** package, which includes a Kerberized ftp client library for Python.

See section 2.3 *Installing Kftp* for installation information. To use the product in a UPS environment, first run:

```
% setup gsspy_krb; setup kftp
```

Then run the **kftpcp** command to copy one or more files to/from Enstore via the Kerberized dCache door. This command can be used from the shell or in a script.

6.2.1 Syntax and Options

```
% kftpcp [<options>] <source_file> <destination_file>
```

The available options include:

-p <port>	ftp server port number
-m <a|p>	ftp server mode; active (default), or passive
-v	verbose mode

Notes:

- If your login id is the same on fndca and your local system, and if they match your Kerberos principal, you can leave off **<your_fndca_login_id>@** in front of **fndca:**.
- Specify the path to the remote file starting from the directory under your /pnfs/<storage_group>/ area. E.g., to specify the remote file /pnfs/my_storage_group/path/to/file on the command line, enter only /path/to/file, including the initial slash.
- If you are authorized for multiple storage groups, specify the path to the remote file starting from the storage group you want to use. E.g., to specify the remote file /pnfs/storage_group_1/path/to/file on the command line, enter /storage_group_1/path/to/file, including the initial slash.

6.2.2 Download a File

To download a stored data file from dCache, run:

```
% kftpcp -p 24127 -m p [-v] \
  [<your_fndca_login_id>@]fndca:</path/to/remote_file> \
  </path/to/local_file>
```

6.2.3 </path/to/remote_file>Upload a File

To upload a new data file, run:

```
% kftpcp -p 24127 -m p [-v] \
  </path/to/local_file> \
  [<your_fndca_login_id>@]fndca:</path/to/remote_file>
```

6.2.4 Examples

To read (download) the stored file `/pnfs/storage_group/mydir/myfile` into a local file of the same name, run:

```
% setup kftp
% kftpcp -p 24127 -m p -v myloginid@fndca:/mydir/myfile \
  /path/to/myfile
    Transferred 42 bytes
```

Or, if your usernames and principal all match, you could run:

```
% kftpcp -p 24127 -m p -v fndca:/mydir/myfile /path/to/myfile
```

Further, if you must specify a particular storage group (`/sg`), run:

```
% kftpcp -p 24127 -m p -v fndca:/sg/mydir/myfile \
  /path/to/myfile
```

6.3 NonKerberized FTP Service (Read-only)

The dCache nonKerberized ftp service currently runs on node the following nodes and corresponding ports:

- fndca.fnal.gov, port 24126 (for STKEN).
- cdfdca, port 25126 (for CDFEN)
- d0endca, port 24126 (for D0EN)

This is read-only, and is not necessarily supported by all experiments. This ftp service can be accessed by ordinary ftp client software. You must specify the host port in your ftp command, as shown below. The Enstore admin will have sent you an email to confirm your registration for this service, and included a password for it.[1] Log in with your username and password.

Sample weakly-authenticated read-only ftp session

Here we explicitly use a nonKerberized ftp client, `/usr/bin/ftp`, and make the connection to fndca port 24126. In the session, we first successfully retrieve a file called `myfile`, and secondly attempt to write a file `trace.txt` and (correctly) fail.

```
% /usr/bin/ftp fndca.fnal.gov 24126
    Connected to stkendca3a.fnal.gov.
    220 FTPDoorIM+PWD ready (read-only server)
    Name (fndca:aheavey):
```

1. If you need to change this password, send email to enstore-admin@fnal.gov.

```
331 Password required for aheavey.
Password: (password entered here)
230 User aheavey logged in
ftp> cd aheavey/test3
250 CWD command succcessful. New CWD is </aheavey/test3>
ftp> ls
200 PORT command successful
150 Opening ASCII data connection for file list
myfile
myfile2
myfile3
226 ASCII transfer complete
10 bytes received in 0.018 seconds (0.55 Kbytes/s)
ftp> get myfile
200 PORT command successful
150 Opening BINARY data connection for
  /pnfs/fs/usr/test/aheavey/test3/myfile
226 Closing data connection, transfer successful
local: myfile remote: myfile
42 bytes received in 0.05 seconds (0.82 Kbytes/s)
ftp> put trace.txt
200 PORT command successful
500 Command disabled
ftp> bye
```

6.4 Grid (GSI) FTP

GSI FTP uses Grid Proxies for authentication and authorization and is compatible with popular tools such as globus-url-copy (from the globus toolkit available at `http://www.globus.org` or from sam_gridftp in Kits). GSI FTP currently runs on the following nodes and corresponding ports:

- fndca.fnal.gov, port 2811 (for STKEN)
- cdfdca, port 2811 (for CDFEN)
- d0endca, port 2811 (for D0EN)

It is more convenient to run this through an interface like srmcp (see section 6.4.3 *Storage Resource Management (SRM)*) which allows you to perform multiple transfers in a single command, and in addition optimizes the parameters of the transfer.

6.4.1 Obtaining Grid Proxies

Grid proxies can be issued automatically for Fermilab users authenticated to Kerberos. See `http://computing.fnal.gov/security/pki/` for instructions.

For non-Fermilab people, Grid proxies can be created from X509 certificates issued by DOE science grid. Install the globus toolkit from `http://www.globus.org`, then run:

```
% grid-proxy-init
```

Then to transfer files, you run the **`globus-url-copy`** command, for example:

```
% globus-url-copy gsiftp://fncda.fnal.gov:2811/<pnfs_path>
  file:///<local_path>
```

This transfers a file from Enstore to a local disk. You can copy in the other direction, e.g.,

```
% globus-url-copy file:///<local_path>
  gsiftp://fncda.fnal.gov:2811/<pnfs_path>
```

and you can copy from one Enstore system to another, e.g., from STKEN to D0EN.

6.4.2 GSI FTP with Kftpcp

GSI FTP can be used with kftpcp (see section 6.2 *Kerberized ftp via the kftpcp Command*), either with Grid proxies or Kerberos tickets. Install and setup kftp (from Kits). Then install and setup either gsspy_gsi (for Grid proxy) or gsspy_krb (for Kerberos); both products are in Kits.

6.4.3 Storage Resource Management (SRM)

SRM is the middleware for managing storage resources for the grid. The SRM implementation within the dCache manages the dCache/Enstore system. It provides functions for file staging and pinning, transfer protocol negotiation and transfer url resolution.

The SRM client "srmcp" provides a convenient way to transfer multiple files from/to dCache/Enstore using variety of protocols. An additional benefit of srmcp is the optimization of tape access in the case of multiple file reads.

To read about SRM, go to `http://sdm.lbl.gov/`, click on **PROJECTS**, and look for *Storage Resource Management (SRM) Middleware Project*.

Srmcp is the implementation of srm client as specified by srm spec v1.0 `http://sdm.lbl.gov/srm/documents/joint.docs/srm.v1.0.doc`. You can use srmcp for optimizing the retrieval and/or storage of files to/from Mass Storage Systems (MSS) which implement SRM. In this document we are interested in using SRM to get/put files from/to Fermilab Enstore via dCache.

Install java on your system.

In order to use Grid Security Interface with srmcp, follow the instructions in the README.SECURITY file that comes with srmcp v1_2 in Kits.

Command Syntax

```
% java SRMCopy [command line options] source(s) destination
```

Default options will be read from a configuration file but can be overridden by the commnad line options. The options are listed and defined in the srmcp v1_2 README file in Kits. We do not list them here.

6.4.4 Examples

These examples are taken from the srmcp v1_2 README file in Kits (with unnecessary options removed).

The following command will retrieve two files `/mypath/myfile1.ext` and `/mypath/myfile1.ext` from enstore via dCache and store them in the local directory `/home/timur/targetdir`:

```
% srmcp \
  srm://cdfendca3.fnal.gov:25129//mypath/myfile1.ext \
  srm://cdfendca3.fnal.gov:25129//mypath/myfile1.ext \
  file://localhost//home/timur/targetdir
```

The following will copy the same files from one storage to another:

```
% srmcp \
  srm://cdfendca3.fnal.gov:25129//mypath/myfile1.ext \
  srm://cdfendca3.fnal.gov:25129//mypath/myfile1.ext \
  srm:/stkendca.fnal.gov:24128/targetdir
```

The following will get the file using dccp client (it should be already installed on you machine)

```
% srmcp  \
  -protocols=dcap   \
  srm:/stkendca.fnal.gov:24128/targetdir/myfile1.ext
  file:////tmp/myfile1.ext
```